



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Generalized linear mixed models: an application in fungal occurrence data

Citation for published version:

Oliveira, TDP, Moral, RA, Hinde, J, Demetrio, CGB & Zocchi, SS 2015, 'Generalized linear mixed models: an application in fungal occurrence data', 60 Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria e 16 Simpósio de Estatística Aplicada à Experimentação Agronômica, Presidente Prudente, Brazil, 8/06/15 - 14/06/15.

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Publisher's PDF, also known as Version of record

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Generalized linear mixed models: an application in fungal occurrence data

Thiago P. Oliveira¹, Rafael A. Moral¹, John Hinde², Clarice G.B. Demétrio¹, Silvio S. Zocchi¹

¹ Universidade de São Paulo, Piracicaba, Brazil

² National University of Ireland Galway, Galway, Ireland

E-mail for correspondence: thiago.paula.oliveira@gmail.com

Abstract: When analysing proportion data, a useful framework is that of generalized linear models. Random effects may be included in the linear predictor for different reasons, e.g., to incorporate correlation between observations taken within the same subject or to model overdispersion. In this work, we use binomial mixed models to model the occurrence of entomopathogenic fungi in five different Brazilian biomes in the dry and humid seasons of 2012. We add an observation-level random effect to incorporate overdispersion and test for the significance of the interaction effect between biome and season.

Keywords: Overdispersion, Biological control; `lme4` package; Logistic-normal model.

1 Introduction

In Brazil there are several ecosystems and living organisms that are associated to the climatic conditions, soil, water and other factors. A range of microorganisms live in the soil, amongst them the fungi, which may contribute to arthropod regulation. There are several products made with fungi used to control insects in different crops, e.g., *Beauveria bassiana*, *Isaria fumosorosea* and *Metarhizium anisopliae*, see Faria and Wraight (2007). In this context, it is important to understand the occurrence of these entomopathogenic fungi in different Brazilian biomes to plan preservation strategies and test their potential in controlling different pests.

This paper was published as a part of the proceedings of the 30th International Workshop on Statistical Modelling, Johannes Kepler Universität Linz, 6–10 July 2015. The copyright remains with the author(s). Permission to reproduce or extract any parts of this abstract should be requested from the author(s).

2 Material and methods

2.1 Case-study

To study the occurrence of entomopathogenic fungi, soil samples were collected in areas that represented the native vegetation of the Brazilian biomes Amazon, Caatinga, Atlantic Forest, Cerrado and the Pampas in the humid and dry seasons of 2012. Sampling was made in four different farms per biome and in each farm, six points were sampled, totaling 120 observations per season. Then, 10 larvae of *Galleria mellonella* and 10 larvae of *Tenebrio molitor* were exposed to each soil sample and after three weeks, the number of insects dead due to fungal infection was observed. This is a hierarchical design and observations within the same farm are correlated so this must be taken into account in the modelling process.

2.2 Statistical models

The data in this experiment consist of proportions so a reasonable assumption is that the number of insects dead due to fungal infection $Y_{ijk} \sim \text{Bin}(m_{ijk}, \pi_{ijk})$. An initial step was to fit a binomial generalized linear mixed model with logit link and the linear predictor

$$\text{logit}(\pi_{ijk}) = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \sigma_Z z_{ij}, \quad (1)$$

where μ is the intercept, α_i is the effect of the i -th biome, γ_j is the effect of the j -th season, $(\alpha\gamma)_{ij}$ is the interaction between the i -th biome and the j -th season, σ_Z^2 is the variance of the random effect associated with the farms and Z_{ij} are standard normal random variables.

To model overdispersion, a random effect at the observation level was included in the linear predictor, so that model (1) became

$$\text{logit}(\pi_{ijk}) = \mu + \alpha_i + \gamma_j + (\alpha\gamma)_{ij} + \sigma_Z z_{ij} + \sigma_W w_{ijk}, \quad (2)$$

where σ_W^2 is the variance of the random effect at the observation level and W_{ijk} are standard normal random variables. This model is also called a logistic-normal model, see Demétrio et al. (2014).

To test for the significance of fixed effects, we used likelihood-ratio tests for nested models. Goodness-of-fit was assessed via half-normal plots with simulation envelopes (Moral et al., 2014) using R (R Core Team, 2014).

3 Results and discussion

We verified that for both species model (1) did not fit well to the data, see Figures 1(a) and (c). So we studied the inclusion of an observation-level random effect (model (2)) to incorporate overdispersion, resulting in a better model fit, see Figures 1(b) and (d). The interaction between biome

and season was significant at 5% significance level for both species, see Table 1 for estimated parameters. It was found that a larger proportion of entomopathogenic fungi infected *G. mellonella* in the dry season when compared to the humid, however the opposite result was obtained for *T. molitor* in the Amazon, see Figure 2. Moreover, there is a clear difference between the proportion of infected insects for humid and dry seasons in the Caatinga biome for *G. mellonella*, which was not found for *T. molitor*. Apparently, the patterns are the same for both species in the biomes Cerrado, Atlantic Forest and Pampas. The data is zero-inflated for some biome \times season combinations and this is subject of ongoing work.

TABLE 1. Parameter estimates for model (2) fitted to both species' data.

Parameter	Species	
	<i>G. mellonella</i>	<i>T. molitor</i>
Intercept (μ)	-0.28 (0.30)	-3.88 (0.49)
Caatinga (α_2)	1.24 (0.43)	1.09 (0.62)
Cerrado (α_3)	-1.89 (0.46)	1.46 (0.61)
Atlantic Forest (α_4)	-1.66 (0.46)	-0.15 (0.69)
Pampas (α_5)	-1.92 (0.46)	4.92 (0.61)
Humid season (γ_2)	-3.34 (0.55)	1.78 (0.60)
Caatinga \times humid season ($\alpha\gamma_{22}$)	3.74 (1.21)	-1.42 (0.82)
Cerrado \times humid season ($\alpha\gamma_{32}$)	5.96 (0.73)	0.61 (0.78)
Atlantic Forest \times humid season ($\alpha\gamma_{42}$)	1.75 (0.77)	-2.53 (0.99)
Pampas \times humid season ($\alpha\gamma_{52}$)	3.81 (0.73)	-1.99 (0.77)
σ_Z	0.0143	0.0898
σ_W	1.2766	1.4082

Acknowledgments: Special Thanks to CNPq and FAPESP.

References

- Demétrio, C.G.B., Hinde, J. and Moral, R.A. (2014). Models for overdispersed data in entomology. In: Ferreira, C.P. and Godoy, W.A.C. (Eds.) *Ecological Modelling Applied to Entomology*, Springer, 219–259.
- Faria, M.R. and Wraight, S.P. (2007). Mycoinsecticides and mycoacaricides: A comprehensive list with worldwide coverage and international classification of formulation types. *Biological Control*, **43**, 237–256.
- Moral, R.A., Hinde, J. and Demétrio, C.G.B. (2014). hnp: Half-normal plots

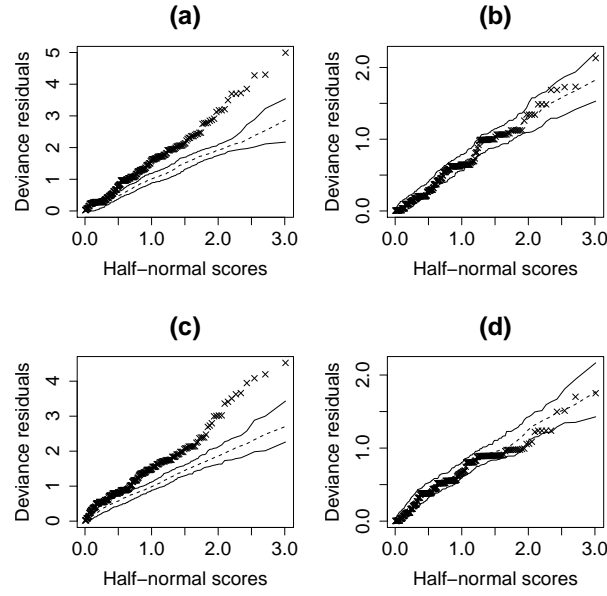


FIGURE 1. Half-normal plots with simulation envelopes for (a) model (1) and (b) model (2) for *G. mellonella* and (c) model (1) and (d) model (2) for *T. molitor*.

with simulation envelopes. R package version 1.0. URL <http://CRAN.R-project.org/package=hnp>.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

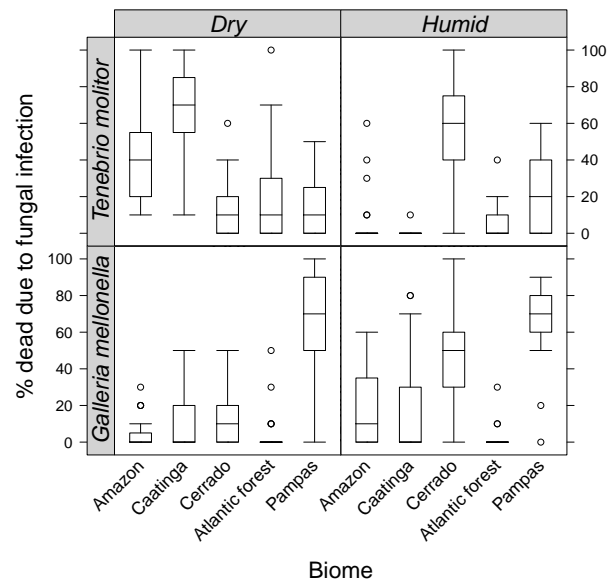


FIGURE 2. Box-plots for each biome×season combination for both species.